

Legitimate Interest as a Legal Basis for Training AI Models

September 2026

When, and under what conditions, can legitimate interest be relied on for the collection and processing of personal data by Artificial Intelligence (AI) systems?

AI systems increasingly process personal data at every stage of their lifecycle, from model training to output generation. Because this processing rarely fits within a consent-based or contract-based legal basis, controllers frequently invoke legitimate interest under Article 6(1)(f) of the General Data Protection Regulation (GDPR).

While there has been debate as to whether legitimate interest is appropriate as a legal basis for the processing of personal data in AI systems, the regulatory consensus has now clarified the matter, undoubtedly influenced by national policy recognising AI as an undeniable competitive priority. But like all things in life, it's not as simple as it looks: relying on legitimate interest as a legal basis is neither automatic nor a mere check-the-box exercise. Drawing on the recommendations from the European Data Protection Board (EDPB)¹, EU supervisory authorities expect data controllers to conduct appropriate due diligence, which can be burdensome in practice, when relying on legitimate interest (more on that below).

Legitimate interest? Yes, but it depends...

Article 6(1)(f) GDPR permits processing that is necessary for the purposes of legitimate interests pursued by the controller or a third party, *provided* those interests are not overridden by the interests or fundamental rights of the data subject. The EDPB's Guidelines 1/2024, applied to AI in Opinion 28/2024, breaks this down into three cumulative conditions,² each of which must be independently satisfied.

1. Real, specific and lawful?

The interest must be *real, specific and lawful*, not speculative or vaguely framed. "*Improving our AI models*" or "*AI innovation*", standing alone, will not satisfy a supervisory authority; the interest must be articulated with reference to a concrete business purpose (e.g. detecting fraudulent transactions,

¹EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, adopted 17 December 2024, para. 40 et seq., confirming that legitimate interest may in principle serve as the legal basis for both the development and deployment phases of an AI model.

²EDPB, Guidelines 1/2024 on the processing of personal data based on Article 6(1)(f) GDPR, adopted 8 October 2024; applied to AI models in EDPB Opinion 28/2024, paras 43–44.

training a specific classifier, or improving the accuracy of a named product feature). The interest pursued should not be confused with the purpose of the processing. Purpose is the specific reason data is processed; interest is the *broader benefit the controller seeks*.

2. Really necessary?

The processing must be the least intrusive means reasonably available to achieve that interest. Where anonymisation, synthetic data, data minimisation, filtering, or a smaller training corpus could reasonably achieve the same purpose, reliance on personal data (and therefore on legitimate interest) becomes harder to sustain. This mirrors the European Court of Justice's (CJEU) reasoning in *Meta Platforms and Others v Bundeskartellamt*, where necessity was assessed against the availability of “reasonable, just as effective, but less intrusive alternatives.”³

3. A balancing act

The controller's interest must not be overridden by the data subject's interests, rights and freedoms, assessed in light of their reasonable expectations at the point of collection, the nature and sensitivity of the data, and the potential impact of the processing. In *Meta v Bundeskartellamt*, the CJEU held that the personalised-advertising processing at issue could not be justified under legitimate interest because users could not reasonably have expected the extent of the processing involved, even though the underlying commercial interest was itself legitimate.⁴ The same reasoning transposes directly to AI: the more opaque or unexpected the processing (e.g. large-scale scraping of online content for model training), the less likely the balancing test resolves in the controller's favour.

Not without risks

Several possible scenarios illustrate where relying on legitimate interest can fail in an AI context:

- Bulk collection of publicly available data for model training does not automatically pass the balancing test; individuals rarely expect content they posted for one purpose to be repurposed to train a commercial model.
- The EDPB treats each phase of the AI lifecycle as a distinct processing operation requiring its own legal basis assessment; a valid basis for training does not automatically extend to deployment-time processing.⁵
- Legitimate interest cannot independently authorise special category data; an Article 9(2) exception is required.
- Where AI output produces legal or similarly significant effects on the individual, Article 22 of the GDPR applies in parallel and legitimate interest alone will rarely be sufficient without additional safeguards.

Not as easy as it looks

In practice, controllers relying on legitimate interest for AI processing should prepare a Legitimate Interest Analysis (LIA) before processing begins, and repeat that exercise for each new AI use case built on existing data rather than extending a prior LIA by analogy. A use case that appears similar on its face may raise materially different necessity or balancing considerations. The necessity analysis should

³ CJEU, Case C-252/21, *Meta Platforms Inc. and Others v Bundeskartellamt*, judgment of 4 July 2023, paras 108–109 and 121.

⁴ CJEU, Case C-252/21 (n 3), and commentary in Herbert Smith Freehills Kramer, “CJEU: Commercial interest may be considered necessary for controller's legitimate interest” (2024), noting the balancing test factors under Recital 47 GDPR, including reasonable expectations and the nature of the controller-data subject relationship.

⁵ EDPB, Opinion 28/2024 (n 1), paras 20–23, distinguishing the “development” phase (data collection, pre-processing, training) from the “deployment” phase (operational use of the trained model), and requiring separate legal basis analysis for each.

be documented explicitly, identifying and ruling out less intrusive alternatives such as pseudonymisation, filtering, synthetic data, or smaller datasets, rather than asserting necessity in general terms that a supervisory authority would be unlikely to accept.

Controllers should also provide an accessible, easy-to-execute means of exercising the right to object under Article 21 of the GDPR and honour it promptly, since an opt-out mechanism that is unused in practice or difficult to exercise tends to weaken the outcome of the balancing test rather than support it. A data-lifecycle map that separates training-data collection, fine-tuning, and deployment (each with its own documented legal basis) is similarly important, given that these are treated as distinct processing operations rather than a single undifferentiated activity.

Finally, the LIA should be reassessed whenever the AI system's purpose, scale, data sources, or risk profile changes materially, and before any new deployment of an existing model, treating the assessment as a living document rather than a one-off exercise performed at launch.

Closing thoughts

Legitimate interest remains, in many cases, the most realistic legal basis for AI-driven processing, but it is a substantive, evidence-based test, not a default. Both the EDPB and the CJEU now expect controllers to demonstrate a genuine, case-specific balancing exercise, reinforced by concrete technical and organisational safeguards. It is therefore recommended to conduct a legal basis review at the design stage of any new AI initiative, and before scaling an existing one.